Int. J. on Recent Trends in Engineering and Technology, Vol. 9, No. 1, July 2013

Reliability Demonstration Testing for Discrete-Type Software Products Based on Variation Distance

Kiyoshi Sawada
Department of Policy Studies,
University of Marketing and Distribution Sciences, Kobe, Japan
Email: Kiyoshi_Sawada@red.umds.ac.jp

***Abstract*—Reliability demonstration testing for software products is performed for the purpose of examining whether the specified reliability is realized in the software after the development process is completed. This study proposes a model of reliability demonstration testing for discrete-type software such as software for numerical calculations. The number of input data sets for test and acceptance number of input data sets causing software failures in the test are designed based on variation distance. This model has less parameters to be prespecified than the statistical model.**

***Index Terms*—software reliability, reliability demonstration testing, discrete-type software, variation distance**

I. Introduction

Many studies on estimating software reliability analyze data obtained in the development process of a software product [1, 2]. On the other hand, reliability demonstration testing which has been developed for hardware products originally [3] should also be applied to software products for the purpose of examining whether the specified reliability has been attained after the development process [4].

We have suggested models for reliability demonstration testing of discrete-type software such as software for numerical calculations [5]. The number of input data sets for test and acceptance number of input data sets causing software failures in the test are designed based on the concept of a statistical test, which requires us to specify the values of producer's and consumer's risks. This study proposes a model of reliability demonstration testing for discrete-type software on the basis of variation distance [6], which can be regarded as a measure to express the distance between two probability distributions. This model can design the test more easily than the statistical model since this model includes less parameters to be prespecified than the statistical model.

II. Notation and Assumptions

This study discusses software reliability demonstration testing (SRDT) model for software which is used discretely in time. It is convenient to evaluate the reliability of this type of software in terms of the probability, p that a software failure occurs for an arbitrarily selected input data set. This probability is called *unreliability* of the software in the following.

This study considers SRDT, where the software of interest is tested with n input data sets, and is accepted if the number of input data sets causing software failures in the test does not exceed an integer c and rejected otherwise. In this case, the design variables are n ($n = 1, 2, \ldots$) and c ($c = 0, 1, 2, \ldots, n-1$).

The notation used in this paper is as follows:

p_0 Unreliability of the software on the contract

p_1 Tolerable upper limit for unreliability of the software ($p_0 < p_1$)

E_1 Event that the number of input data sets causing software failures in the test exceeds c

$\overline{E}_1$ Complement of E_1

The assumptions made throughout this paper are listed below:

(i) No fault is removed during the test. All the faults for software failures during the test are removed after the test is completed.

(ii) The values of p_0 and p_1 are specified at the beginning of the test.

III. Statistical Model

For the purpose of determining the values of n and c in the above SRDT, this section presents a model based on the concept of a statistical test.

In the above SRDT, the probability that event E_1 occurs when $p = p_0$ is given by

$$\Pr[E_1 \mid p_0] = \sum_{i=c+1}^{n} \binom{n}{i} p_0^{\,i} (1-p_0)^{n-i}. \tag{1}$$

Likewise, the probability that event $\overline{E}_1$ occurs when $p = p_1$ is expressed by

$$\Pr[\overline{E}_1 \mid p_1] = \sum_{i=0}^{c} \binom{n}{i} p_1^{\,i} (1-p_1)^{n-i}. \tag{2}$$

The probability in Eq. (1) is called a *producer's risk* in SRDT as well as in sampling theory, while that in Eq. (2) is called a *consumer's risk*. The producer's and consumer's risks, respectively, signify the probabilities of Type I and Type II errors in terms of a statistical test.

When the values of the producer's and consumer's risks are specified to be equal to or less than α and β, respectively, a feasible region for designing a software reliability demonstration test is written as

$$\sum_{i=c+1}^{n} \binom{n}{i} p_0^{\,i} (1-p_0)^{n-i} \le \alpha, \tag{3}$$

DOI: 01.IJRTET.9.1.45

Int. J. on Recent Trends in Engineering and Technology, Vol. 9, No. 1, July 2013

$$\sum_{i=0}^{c}\binom{n}{i}p_1^{\,i}(1-p_1)^{n-i}\le\beta. \tag{4}$$

In general, the simultaneous equations obtained by replacing inequalities in Eqs. (3) and (4) by equalities do not always have a unique solution since *n* and *c* are integers. A practical solution would be obtained by using the following four conditions:

(i) The producer's risk does not exceed α.

(ii) The consumer's risk does not exceed β.

(iii) The number of input data sets *n* is the minimum.

(iv) Acceptance number *c* is the minimum.

IV. Variation Distance Model

This section presents a model based on the variation distance with a view to determining the values of *n* and *c* in the above SRDT.

A. Formulation of Variation Distance

Let F_1 and F_2 denote two types of discrete probability distribution, and q_{1i} and q_{2i} (i = 0, 1, 2, ...) be probability mass functions associated with F_1 and F_2, respectively. Then the variation distance [6] of F_1 and F_2 is defined by

$$d_v(F_1,F_2)\equiv\sum_{i=0}^{\infty}|\,q_{1i}-q_{2i}\,|. \tag{5}$$

The variation distance in Eq. (5) can be regarded as a measure to express the distance between F_1 and F_2. In other words, as the variation distance increases, we can distinguish F_1 from F_2 more easily.

When $p = p_0$, we have

$$\Pr[E_1\mid p_0]=\sum_{i=c+1}^{n}\binom{n}{i}p_0^{\,i}(1-p_0)^{n-i}, \tag{6}$$

$$\Pr[\overline{E}_1\mid p_0]=\sum_{i=0}^{c}\binom{n}{i}p_0^{\,i}(1-p_0)^{n-i}. \tag{7}$$

In the case of $p = p_1$, we have

$$\Pr[E_1\mid p_1]=\sum_{i=c+1}^{n}\binom{n}{i}p_1^{\,i}(1-p_1)^{n-i}, \tag{8}$$

$$\Pr[\overline{E}_1\mid p_1]=\sum_{i=0}^{c}\binom{n}{i}p_1^{\,i}(1-p_1)^{n-i}. \tag{9}$$

A set of Eqs. (6) and (7) represents a probability distribution F_1, expressed by a probability mass function q_{1i} (i = 0, 1) in Eq. (5) when $p = p_0$. A set of Eqs. (8) and (9) expresses another probability distribution F_2, characterized by a probability mass function q_{2i} (i = 0, 1) in Eq. (5) in the case of $p = p_1$. Hence, the variation distance of these two probability distributions is written as

$$d_v(F_1,F_2)=2\left[\sum_{i=0}^{c}\binom{n}{i}p_0^{\,i}(1-p_0)^{n-i}-\sum_{i=0}^{c}\binom{n}{i}p_1^{\,i}(1-p_1)^{n-i}\right]. \tag{10}$$

We can obtain an optimal pair of values, $(n, c)^*$ by maximizing $d_v(F_1, F_2)$ in Eq. (10) since we can distinguish F_1 from F_2 more easily as the variation distance increases.

B. An Optimal Value c^* for Each Value of n

This section shows the existence of the optimal value c^* for each value of *n*.

Let $D(c) = d_v(F_1, F_2)$, the difference of $D(c)$ is given by

$$D(c+1)-D(c)=2\binom{n}{c+1}\left[p_0^{\,c+1}(1-p_0)^{n-c-1}-p_1^{\,c+1}(1-p_1)^{n-c-1}\right]. \tag{11}$$

The sign of $D(c + 1) – D(c)$ follows that

(i) If

$$c<\frac{n\log\dfrac{1-p_1}{1-p_0}}{\log\dfrac{p_0}{p_1}+\log\dfrac{1-p_1}{1-p_0}}-1, \tag{12}$$

then $D(c+1)-D(c)>0$.

(ii) If

$$c=\frac{n\log\dfrac{1-p_1}{1-p_0}}{\log\dfrac{p_0}{p_1}+\log\dfrac{1-p_1}{1-p_0}}-1, \tag{13}$$

then $D(c+1)-D(c)=0$.

(iii)If

$$c>\frac{n\log\dfrac{1-p_1}{1-p_0}}{\log\dfrac{p_0}{p_1}+\log\dfrac{1-p_1}{1-p_0}}-1, \tag{14}$$

then $D(c+1)-D(c)<0$.

Therefore, Let

$$\psi=\frac{n\log\dfrac{1-p_1}{1-p_0}}{\log\dfrac{p_0}{p_1}+\log\dfrac{1-p_1}{1-p_0}}-1, \tag{15}$$

then we have the following theorem.

Theorem 1:

(i) If ψ is not integer, there exists a unique c^* that maximizes $D(c)$ and c^* is the minimum integer which is greater than .

(ii) If is integer, there exist two c^* that maximizes $D(c)$ and c^* are and + 1.

DOI: 01.IJRTET.9.1.45

V. Relation Between Statistical Model And Variation Distance Model

This section reveals relation between the statistical model in Section III and the variation distance model in Section IV. $d_v(F_1, F_2)$ in Eq. (10) is expressed as following:

$$d_v(F_1, F_2) = 2 - 2\left(\Pr[E_1 \mid p_0] + \Pr[\overline{E}_1 \mid p_1]\right) \qquad (16)$$

using the producer's risk $\Pr[E_1 \mid p_0]$ of Eq. (1) and the consumer's risk $\Pr[\overline{E}_1 \mid p_1]$ of Eq. (2).

Therefore, maximizing the variation distance $d_v(F_1, F_2)$ is equivalent to minimizing the sum of the producer's risk $\Pr[E_1 \mid p_0]$ and the consumer's risk $\Pr[\overline{E}_1 \mid p_1]$ in the statistical model. This means to minimize the expected cost in case the cost of the producer's risk is equal to the cost of the consumer's risk.

VI. Numerical Examples

Figure 1 shows numerical examples of variation distance $d_v(F_1, F_2)$ when acceptance number of input data sets causing software failures c varies in $c = 0, 1, 2, ..., 10$ for $n = 1000, 2000, 3000, 4000$, and 5000 where $p_0 = 0.001$ and $p_1 = 0.002$.

For each value of $n = 1000, 2000, 3000, 4000$, and 5000, the optimal values are $c^* = 1, 2, 4, 5$, and 7, respectively. It is observed that c^* increases and its corresponding variation distance increases with increasing n.

VII. Conclusions

This study proposed a model of reliability demonstration testing for discrete-type software such as software for numerical calculations. When unreliability of the software on the contract p_0 and tolerable upper limit for unreliability of the software p_1 are given, the number of input data sets for test n and acceptance number of input data sets causing software failures in the test c are designed by maximizing $d_v(F_1, F_2)$. This model has less parameters to be prespecified than the statistical model. Theorem 1 reveals the existence of the optimal value c^* for each value of n.

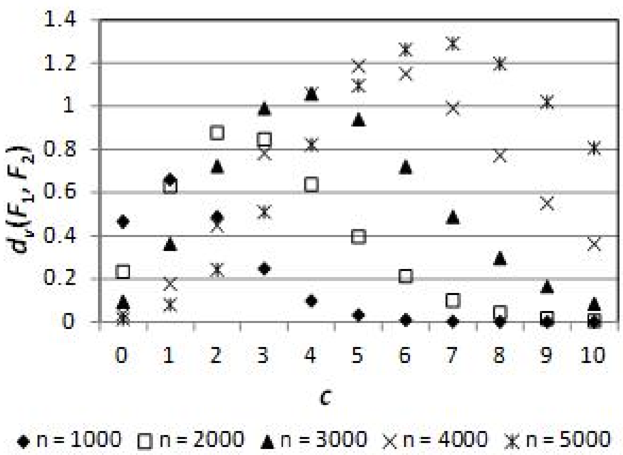

Figure 1. Numerical examples of variation distance

References

[1] H. Pham, System Software Reliability, London: Springer, 2006.

[2] N. D. Singpurwalla and S. P. Wilson, Statistical Methods in Software Engineering: Reliability and Risk, New York: Springer, 1999.

[3] N. R. Mann, R. E. Schafer, and N. D. Singpurwalla, Methods for Statistical Analysis of Reliability and Life Data, New York: John Wiley & Sons, pp.404–410, 1974.

[4] H. Sandoh, "Reliability demonstration testing for software", IEEE Transactions on Reliability, vol.40, pp.117–119, 1991.

[5] K. Sawada and H. Sandoh, "A summary of software reliability demonstration testing models", International Journal of Reliability, Quality and Safety Engineering, vol.6, pp.65–80, 1999.

[6] S. Amari, et al., Differential Geometry in Statistical Inference, Institute of Mathematical Statistics, 1987.

DOI: 01.IJRTET.9.1.45